Magic state distillation with punctured polar codes

Anirudh Krishna, Jean-Pierre Tillich

Magic state distillation with punctured polar codes

Anirudh Krishna[1] and Jean-Pierre Tillich[2]

[1] *Université de Sherbrooke, 2500 Boulevard de l'Université, Sherbrooke, QC J1K 2R1, Canada*
[2] *Inria, Team SECRET, 2 Rue Simone IFF, CS 42112, 75589 Paris Cedex 12, France*

Abstract

We present a scheme for magic state distillation using punctured polar codes. Our results build on some recent work by Bardet et al. [1] who discovered that polar codes can be described algebraically as decreasing monomial codes. Using this powerful framework, we construct tri-orthogonal codes [2] that can be used to distill magic states for the T gate. An advantage of these codes is that they permit the use of the successive cancellation decoder whose time complexity scales as $O(N\log(N))$. We supplement this with numerical simulations for the erasure channel and dephasing channel. We obtain estimates for the dimensions and error rates for the resulting codes for block sizes up to 2^{20} for the erasure channel and 2^{16} for the dephasing channel. The dimension of the triply-even codes we obtain is shown to scale like $O(N^{0.8})$ for the binary erasure channel at noise rate 0.01 and $O(N^{0.84})$ for the dephasing channel at noise rate 0.001. The corresponding bit error rates drop to roughly 8×10^{-28} for the erasure channel and 7×10^{-15} for the dephasing channel respectively.

1 Introduction

Implementations of quantum circuits are imperfect and prone to error. In order to realize scalable quantum computers, we need to construct quantum circuits capable of working with unreliable components. This is the focus of the domain of fault-tolerant quantum computation [3, 4, 5, 6]. The premise behind this theory is to encode quantum information using *quantum error correcting codes* which serve as a buffer against noise.

To process encoded information, we require some way to perform logical operations without unencoding it and thereby leaving it vulnerable to errors. Of particular interest is the technique called state injection, a scheme to inject special ancilla states called magic states into a quantum circuit [7]. These states must undergo a resource intensive purification process called magic state distillation to ensure that they have high fidelity. Current estimates state that a large fraction of qubits required for quantum computation will have to be dedicated to this process [8, 9]. It is therefore imperative to reduce the overhead required by magic state distillation as this is a bottleneck.

We shall focus on magic state distillation protocols to distill the state $|A\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$. This state can be used to inject the T gate, where

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} . \tag{1}$$

Together with gates from the Clifford group, this gate can be used to implement a universal set of gates. There are several magic state distillation protocols to distill the $|A\rangle$ state such as [7, 10, 11] that use quantum error correcting codes such as the Reed-Muller code with special symmetry properties. These symmetries imply that applying transversal, physical T gates to the physical qubits of the quantum error correcting code implements the logical T gate on the encoded qubits. Decoding the quantum error correcting code then results in the magic state $|A\rangle$. The framework of tri-orthogonality defined by Bravyi and Haah [2] extends these symmetries to describe entire code families. This work also defined a protocol to distill states using tri-orthogonal codes. One key feature of the Bravyi-Haah protocol is that it relies on post-selection, i.e. the

states at the end of the protocol are accepted if and only if a certain measurement outcome is obtained; if not, the states are discarded. The probability of obtaining this measurement outcome decreases as the distance of the quantum error correcting code increases.

In this paper, we present a technique to distill the state $|A\rangle$ using punctured polar codes. We shall show that these codes are tri-orthogonal and can encode a growing number of logical qubits similar to [10]. Rather than use post selection, we modify the Bravyi-Haah protocol and use Steane error correction to correct errors and decode our code. This implies a trade-off between the rate of distillation and the error rate the scheme can tolerate.

Discovered by Arikan [12], polar codes were the first family of codes that were shown to efficiently achieve the capacity of a binary discrete memoryless channel (B-DMC). These codes come equipped with an iterative decoder called the successive cancellation (SC) decoder whose decoding complexity scales as $O(N\log(N))$ for a code of block-size N. We build on some recent results by Bardet et al. wherein polar codes are cast as *decreasing monomial codes* [1]. These algebraic tools are central to demonstrating the existence of tri-orthogonal polar codes. We hope these tools will be useful in the study of quantum error correcting codes in general.

Notably, our magic state distillation has a low decoding complexity as the quantum codes inherit the SC decoder whose time complexity is $O(N\log(N))$. We supplement our analysis with some numerical results which indicate upper bounds on the size of the resulting codes and their error rates for block sizes of interest.

Although Reed-Muller codes are closely related to polar codes, the SC decoder for polar codes has a much better error correction capacity than the majority logic decoding algorithm for Reed-Muller codes [13, 14]. Successive and iterative decoders exist for the Reed-Muller code as well, but again correct far fewer errors than their polar code counterparts [15]. For this reason, using the distance of the polar code to characterize its error correction capacity can be misleading.

Related work: Polar codes have been generalized to the quantum realm [16, 17, 18, 19, 20, 21, 22] but were studied from the perspective of quantum Shannon theory. These authors were not concerned with fault tolerance and thus did not explore transversal gates on quantum codes.

There also exist schemes to distill states to implement gates besides the T gate [23, 24, 25, 26] and also complex schemes that combine both distillation and compilation such as [27, 28, 29, 30].

Finally, magic state distillation can be performed using qudits (d dimensional quantum systems) rather than qubits [31, 32, 33, 34]. These schemes have the potential to reduce the overhead associated with magic state disillation.

Outline: This paper is structured as follows. In section 2, we define binary polar codes, and proceed to describe the algebraic formalism discovered by Bardet et al. [1]. For the sake of completeness, we review the tri-orthogonality condition and its application to magic states. In section 3, we discuss our distillation procedure which assumes that the error correcting code is subject to erasure noise or dephasing noise. We present our results in section 4. First, we compute the dimension of the triply-even codes constructed from polar codes and plot this parameter as a function of the block-size in 4.1. Second, we show that it is possible to construct tri-orthogonal codes from polar codes that achieve good performance for erasure channels and for dephasing channels.

2 Background

2.1 Polar codes

We begin by briefly reviewing the theory of polar codes. Throughout the paper, we let $n \in \mathbb{N}$ denote a natural number and $N = 2^n$. Suppose we wish to transmit a message $x \in \mathbb{F}_2^N$ across N copies of a B-DMC $\mathsf{W} : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} = \mathbb{F}_2$ and $\mathcal{Y}$ is some output alphabet. The fundamental problem of coding theory is to deduce the input word x having received a corrupted word $y \in \mathcal{Y}^N$. Arikan's polarization technique reduces

the task of the decoder to an iterative process.

For a given channel W, block size N and error rate $\epsilon \in [0,1]$, the binary polar code $\mathcal{C} = \mathcal{C}(N,\epsilon)$ is specified by a set $\mathcal{A} \subseteq \{1,...,N\}$, where $|\mathcal{A}| = K$ is the dimension of $\mathcal{C}$ and $\epsilon \in [0,1]$ is an upper bound on the bit error rate of the code. We shall discuss how to obtain the set $\mathcal{A}$ shortly and begin by discussing the encoding and decoding process assuming this set has been provided. The information we wish to transmit is stored in a vector $u \in \mathbb{F}_2^N$ where for $a \in \mathcal{A}$, the indices u_a carry information and the other indices are said to be *frozen*, i.e. for $b \in \mathcal{A}^c$, $u_b = 0$. At the core of the encoding process is the 2×2 matrix F defined as

$$F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} . \tag{2}$$

Given an input word $u \in \mathbb{F}_2^N$, the encoder maps it to $x \in \mathbb{F}_2^N$, where $x = F^{\otimes n}u$. This word x is then transmitted across N copies of the noisy channel W resulting in the corrupted word y.

The decoder is provided the set $\mathcal{A}$ ahead of time. Given y, it deduces the bits of u sequentially (from 1 to N) treating the bits that it has not yet decoded as noise. At the a-th iteration, it sees the synthetic channel $\mathsf{W}^{(a)}$ where

$$\mathsf{W}^{(a)}(y, u_1, \cdots u_{a-1}|u_a) = \frac{1}{2^N} \sum_{u_{a+1},...,u_N} \mathsf{W}(y_1|x_1) \cdots \mathsf{W}(y_N|x_N) .$$

It uses this formula to compute the log likelihood ratios λ_a where

$$\lambda_a = \log_2 \frac{\mathsf{W}^{(a)}(y, u_1, \cdots, u_{a-1}|0)}{\mathsf{W}^{(a)}(y, u_1, \cdots, u_{a-1}|1)} .$$

Having computed the a-th log-likelihood ratio, it estimates the a-th bit as

$$u_a = \begin{cases} 0 & \text{if } \lambda_a > 0 \text{ or } a \in \mathcal{A}^c \\ 1 & \text{if } \lambda_a < 0 \end{cases} .$$

Although it is not evident from this presentation, this computation can be performed in $O(N \log(N))$ steps.

It can be shown that the probability of error of the synthetic channels can be upper bounded using the *Bhattacharyya parameter* $\mathcal{B}(\mathsf{W})$ defined for a B-DMC $\mathsf{W} : \mathbb{F}_2 \to \mathcal{Y}$ as

$$\mathcal{B}(\mathsf{W}) := \frac{1}{2} \sum_{y \in \mathcal{Y}} \sqrt{\mathsf{W}(y|0)\mathsf{W}(y|1)} . \tag{3}$$

The Bhattacharyya parameter obeys $0 \leq \mathcal{B}(\mathsf{W}) \leq 1$, with $\mathcal{B}(\mathsf{W}) = 0$ indicating a noiseless channel and $\mathcal{B}(\mathsf{W}) = 1$ indicating a totally unreliable channel.

Exact code construction for a polar code $\mathcal{C} = \mathcal{C}(N,\epsilon)$ whose bit error rate is upper bounded by $\epsilon \in [0,1]$ employs the Bhattacharyya parameters to deduce the set $\mathcal{A}$. For each index $i \in \{1,...,N\}$, we can evaluate $\mathcal{B}(\mathsf{W}^{(a)})$ by transmitting the all zero codeword 0^N across the channel W^N. Given a threshold $\epsilon \in [0,1)$, we can construct the polar code $\mathcal{C}$ by choosing those indices a such that $\mathcal{B}(\mathsf{W}^{(a)}) \leq \epsilon$ to constitute the set $\mathcal{A}$.

Two channels we shall be interested in are the erasure channel and the binary symmetric channel. When a symbol is erased, we replace it with a special symbol $\perp$. For ease of representation, we denote a probability distribution $\{\mathsf{Pr}(0), \mathsf{Pr}(1), \mathsf{Pr}(\perp)\}$ over symbols in $\mathbb{F}_2 \cup \{\perp\}$ as $\mathsf{Pr}(0)[0] + \mathsf{Pr}(1)[1] + \mathsf{Pr}(\perp)[\perp]$.

For $p \in [0,1]$, the action of the erasure channel, henceforth denoted W_p, on $x \in \mathbb{F}_2$ is then

$$\mathsf{W}_p(x) = (1-p)[x] + p[\perp] ,$$

which is equivalent to saying that the conditional probability of obtaining x at the output given that x was transmitted is $1-p$ and the probability of seeing $\perp$ is p. Similarly, for $p \in [0,1]$, the action of the binary symmetric channel, henceforth denoted V_p, on $x \in \mathbb{F}_2$ is then

$$\mathsf{V}_p(x) = (1-p)[x] + p[\bar{x}] ,$$

where $\bar{x} = x + 1 \pmod 2$.

2.2 Decreasing monomial codes

The drawback with the above construction is that it is not universal; the frozen channels have to be recomputed for each B-DMC W. Although optimizing the polar code this way permits us to provably achieve the capacity efficiently, it does not provide any insight into which synthetic channels are used to transmit information and which synthetic channels are frozen. This is also a problem in practice because exact code construction is difficult for most channels. Devising good algorithms for code construction is non-trivial and has been the subject of considerable research [35], [36], [37].

Recently, Bardet et al. [1] discovered that there exists an algebraic framework describing polar codes which sheds some light on which channels carry information and which channels are frozen. We shall exploit this algebraic structure in order to construct tri-orthogonal quantum codes in the following section. Here we shall briefly review relevant concepts from [1] required to discuss magic state distillation. For a complete development, including proofs of claims in this section, we refer the reader to the original paper by Bardet et al. [1].

Let $n \in \mathbb{N}$ be some natural number and $\mathcal{R}_n$ be the ring of polynomials over n variables defined as

$$\mathcal{R}_n = \mathbb{F}_2[x_0, ..., x_{n-1}]/(x_0^2 - x_0, ..., x_{n-1}^2 - x_{n-1}) \ .$$

Any monomial $\alpha \in \mathcal{R}_n$ can be expressed as $x_0^{a_0}...x_{n-1}^{a_{n-1}}$ for some exponents $a_0, ..., a_{n-1} \in \mathbb{F}_2$. The set of all monomials is denoted $\mathcal{M}_n$ where

$$\mathcal{M}_n = \{x_0^{a_0}...x_n^{a_{n-1}} | a_0, ..., a_{n-1} \in \mathbb{F}_2\} \ .$$

Equivalently, we may directly identify the monomial α with the exponents of x_i as $(a_0, ..., a_{n-1})$. The degree of the monomial α is the number of non-zero exponents and is denoted $\mathsf{deg}(\alpha)$.

We can also uniquely identify α with a vector over $\mathbb{F}_2^N$ by evaluating it over the field $\mathbb{F}_2^n$. Define the evaluation map $\mathsf{ev} : \mathcal{R}_n \to \mathbb{F}_2^N$ where

$$\mathsf{ev}(\alpha) = \{\alpha(u)\}_{u \in \mathbb{F}_2^n} \ , \tag{4}$$

which denotes the vector each of whose components is obtained by evaluating α on each of the N points in $\mathbb{F}_2^n$.

Definition 1 (**Monomial code**). *Let $\mathrm{I} \subseteq \mathcal{M}_n$ be a set of monomials in n variables. The corresponding monomial code $\mathcal{C}(\mathrm{I}) \subseteq \mathbb{F}_2^N$ is defined as*

$$\mathcal{C}(\mathrm{I}) := \mathrm{span}\{\mathsf{ev}(\alpha) | \alpha \in \mathrm{I}\} \ . \tag{5}$$

Lemma 2. *For $\mathrm{I} \subseteq \mathcal{M}_n$, the dimension of the code $\mathcal{C}(\mathrm{I})$ is equal to $|\,\mathrm{I}\,|$.*

Proof. Monomials are linearly independent over $\mathcal{R}_n$ and the map ev, being injective, preserves this relation. □

There is a natural partial order on $\mathcal{M}_n$ called the divisibility order defined as follows. For two monomials $\alpha, \beta \in \mathcal{M}_n$ with exponents $\{a_i\}_i$ and $\{b_i\}_i$, we write $\alpha \preccurlyeq_w \beta$ if and only if for all $i \in \{0, ..., n-1\}$,

$$a_i \leq b_i \ . \tag{6}$$

The subscript w indicates a weak order and we can define a strong order as follows.

Definition 3 (**Strong order**). *Let $\alpha, \beta \in \mathcal{M}_n$ be two monomials of the same degree r such that $f = x_{i_1} \cdots x_{i_r}$ and $g = x_{j_1} \cdots x_{j_r}$. We say that $f \preccurlyeq g$ if and only if for all $1 \leq t \leq r$:*

$$i_t \leq j_t \ . \tag{7}$$

We can then extend this definition to include monomials α and β of different degrees via divisibility. If $\mathsf{deg}(\alpha) < \mathsf{deg}(\beta)$, then $\alpha \preccurlyeq \beta$ if and only if there exists a monomial β^* such that $\mathsf{deg}(\alpha) = \mathsf{deg}(\beta^*)$ and $\alpha \preccurlyeq \beta^* \preccurlyeq_w \beta$.

This ordering can apply to an entire set of monomials as follows.

Definition 4 (**Decreasing sets**). *A set* $\mathrm{I} \subset \mathcal{M}_n$ *is* decreasing *if and only if* ($g \in \mathrm{I}$ *and* $f \preccurlyeq g$) *implies that* $f \in \mathrm{I}$.

Furthermore, a code $\mathcal{C}(\mathrm{I})$ *is a decreasing monomial code if* I *is a decreasing set.*

This ordering is important because of the following relation. Given the exponent $(a_0, ..., a_{n-1})$ of a monomial α, let $a = \sum_{t=0}^{n-1} \mathrm{A}_t \, 2^t$ be the integer corresponding to a natural ordering of the exponent.

Lemma 5. *Let* $\alpha, \beta \in \mathcal{M}_n$ *and* a, b *be the integers corresponding to the exponents of* α, β *respectively. We have*

$$\alpha \preccurlyeq \beta \implies \mathcal{B}(\mathsf{W}^{(a)}) \leq \mathcal{B}(\mathsf{W}^{(b)})$$

This in turn implies the central result of Bardet et al. [1] is the following theorem that we state without proof (See theorem 1 of [1]).

Theorem 6. *Polar codes are decreasing monomial codes.*

Recall that a polar code is constructed by ordering the Bhattacharyya parameters associated with the channels $\mathsf{W}^{(a)}$ for $a \in \{1, ..., N\}$. If we choose to include codewords $\mathsf{ev}(\beta)$ for $\beta \in \mathcal{M}_n$ such that $\mathcal{B}(\mathsf{W}^{(b)}) \leq \epsilon$, then we must necessarily have all $\alpha \preccurlyeq \beta$ because of lemma 5. This result is important because it allows for a simple description of the dual of a monomial code.

We denote the complement of a monomial $\alpha \in \mathcal{M}_n$ by $\check{\alpha}$ as

$$\check{\alpha}(x) = x_0^{a_0 \oplus 1} ... x_{n-1}^{a_{n-1} \oplus 1} \ , \tag{8}$$

where $\oplus$ denotes XOR. By extension, for any subset $\mathrm{I} \subseteq \mathcal{M}_n$, the set $\check{\mathrm{I}} = \{\check{\alpha} | \alpha \in \mathrm{I}\}$. Equivalently, the complement $\check{\alpha}$ of α is the smallest monomial such that $\alpha\check{\alpha} = x_0 ... x_{n-1}$, where $x_0 ... x_{n-1}$ is the complete monomial with all the n variables.

We denote the product of two monomials $\alpha, \beta \in \mathcal{M}_n$ as $\alpha \cdot \beta$ where $(\alpha \cdot \beta)(u) = \alpha(u)\beta(u)$ for $u \in \mathbb{F}_2^n$. For two vectors $v, w \in \mathbb{F}_2^N$, we denote the element-wise product as $v * w = (v_1 w_1, ..., v_N w_N)$.

The evaluation map ev maps the element-wise product to the product between monomials, i.e.

$$\mathsf{ev}(\alpha) * \mathsf{ev}(\beta) = \mathsf{ev}(\alpha \cdot \beta) \ . \tag{9}$$

The following lemma is proved in [1] (See Proposition 6).

Lemma 7. *Let* $\mathrm{I} \in \mathcal{M}_n$ *be a decreasing set of monomials and* $\mathcal{C}(\mathrm{I})$ *be the corresponding monomial code. The dual of* $\mathcal{C}(\mathrm{I})$ *is*

$$\mathcal{C}(\mathrm{I})^{\perp} = \mathcal{C}(\mathcal{M}_n \setminus \check{\mathrm{I}}) \ , \tag{10}$$

which is also a decreasing monomial code.

Proof. For $\alpha, \beta \in \mathcal{M}_n$, we may write the inner product $\langle \mathsf{ev}(\alpha), \mathsf{ev}(\beta) \rangle$ as

$$\langle \mathsf{ev}(\alpha), \mathsf{ev}(\beta) \rangle = \sum_{u \in \mathbb{F}_2^n} \alpha(u)\beta(u) \ . \tag{11}$$

Note that by symmetry, the only monomial that is non-zero when summed over $\mathbb{F}_2^n$ is the complete monomial $x_0 ... x_{n-1}$. Therefore the sum in eq. (11) is non-zero if and only if $\check{\alpha} \preccurlyeq \beta$. Therefore if we want the inner product to be 0 for all $\beta \in \mathrm{I}$, then we require $\alpha \in \check{\mathrm{I}}$. This proves the claim that $\mathcal{C}(\mathrm{I})^{\perp} = \mathcal{C}(\mathcal{M}_n \setminus \check{\mathrm{I}})$.

To prove that $\mathcal{C}(\mathcal{M}_n \setminus \check{\mathrm{I}})$ is also a decreasing monomial code, we can show that $\mathcal{M}_n \setminus \check{\mathrm{I}}$ is a decreasing set. Let $\beta \in \mathcal{M}_n \setminus \check{\mathrm{I}}$ and $\alpha \in \mathcal{M}_n$ such that $\alpha \preccurlyeq \beta$. For the sake of contradiction, assume that $\alpha \notin \mathcal{M}_n \setminus \check{\mathrm{I}}$. This in turn means that $\check{\alpha} \in \mathrm{I}$. However, if $\alpha \preccurlyeq \beta$, it implies that $\check{\beta} \preccurlyeq \check{\alpha}$. Since I is a decreasing set 4, this means that $\check{\beta} \in \mathrm{I}$ or that $\beta \in \check{\mathrm{I}}$, which is a contradiction.

This establishes the result. □

2.3 Tri-orthogonal quantum codes

For the sake of completeness, we review the definition of a triply-even space and a tri-orthogonal code as in [2], [11]. These notions are key to understanding quantum error correcting codes which promote physical transversal T gates to logical transversal T gates.

We denote by X and Z the Pauli operators

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Together with the phase $i\mathbb{1}$, these operators generate the Pauli group $\mathcal{P} = \langle i\mathbb{1}, X, Z\rangle$ and can be extended to N qubits as $\mathcal{P}_N = \mathcal{P}^{\otimes N}$. For $a, b \in \mathbb{F}_2^N$, we let $X(a) = \bigotimes_{i=1}^N X^{a_i}$ and $Z(b) = \bigotimes_{j=1}^N Z^{b_j}$. The Pauli group forms the first level of the Clifford hierarchy, $\mathcal{K}^{(1)}$. The Clifford group $\mathcal{K}^{(2)}$ is defined as the set of automorphisms of the Pauli group, i.e.

$$\mathcal{K}^{(2)} = \{U | \forall P \in \mathcal{P}_n : UPU^\dagger \in \mathcal{P}_N\} .$$

This group can be generated by the phase-gate $S = \sqrt{Z}$, the Hadamard gate $H = (X + Z)/\sqrt{2}$ and the controlled-Z gate $\mathsf{cZ} = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes Z$. By themselves, these gates are insufficient to generate a universal gate set. This is rectified by using gates from the third level of the Clifford hierarchy $\mathcal{K}^{(3)}$,

$$\mathcal{K}^{(3)} = \{U | \forall P \in \mathcal{P}_n : UPU^\dagger \in \mathcal{K}^{(2)}\} .$$

Any gate from the third level of the Clifford hierarchy, together with the Clifford group $\mathcal{K}^{(2)}$ is sufficient to generate a universal gate set. One such gate is the T gate, where $T = \mathrm{diag}(1, e^{i\pi/4})$.

CSS codes are a class of quantum error correcting codes for which all the stabilizer generators are either composed entirely of X operators and identity I or entirely of Z operators and identity I [38, 39]. This can be defined using two codes $\mathcal{C}_X, \mathcal{C}_Z \subset \mathbb{F}_2^N$ such that $\mathcal{C}_Z^\perp \subset \mathcal{C}_X$. We obtain a code by mapping the spaces $\mathcal{C}_X^\perp$ and $\mathcal{C}_Z^\perp$ to stabilizer generators of X and Z type respectively, i.e. for $a \in \mathcal{C}_X^\perp$ and $b \in \mathcal{C}_Z^\perp$, we define stabilizer generators $X(a)$ and $Z(b)$. We seek CSS codes for which the T gate applied transversally on the physical level is equivalent to a T gate applied to the logical qubits encoded by the code. To this end, we consider code spaces with the following properties.

Definition 8 (**Triply even space**). *A subspace $\mathcal{V} \subseteq \mathbb{F}^n$ is said to be* triply-even *if for any triple $u, v, w \in \mathcal{V}$,*

$$|u * v * w| = 0 \pmod 2 .$$

Definition 9 (**Tri-orthogonal matrix**). *Let $\mathrm{H} \in \mathbb{F}_2^{m\times n}$ be a matrix whose rows are labelled $\{h^{(a)}\}_{a=1}^m$. We say that* H *is a tri-orthogonal matrix if and only if*

1. *for $1 \le a < b \le m$*

$$|h^{(a)} * h^{(b)}| = 0 \pmod 2$$

2. *for $1 \le a < b < c \le m$,*

$$|h^{(a)} * h^{(b)} * h^{(c)}| = 0 \pmod 2$$

We shall partition H into two matrices H_0 and H_1, where H_0 contains the $(m-k)$ even weight rows of H and H_1 contains the k odd weight rows of H for some $k \in \mathbb{N}$. Let G be the matrix whose rows are orthogonal to H.

To obtain a tri-orthogonal quantum code $\mathsf{CSS}(X, \mathrm{H}_0; Z, \mathrm{G})$ from the tri-orthogonal matrix H, we associate

1. the rows of H_0 with the X stabilizer generators;

2. the rows of G with the Z stabilizer generators;

3. the rows of H_1 to both the X and Z logical operators.

As shown in the paper by Bravyi and Haah, such a code is useful because they promote transversal T gates to logical T gates. Letting $T_n = T^{\otimes n}$, lemma 2 from [2] states that

Lemma 10. *Suppose* $\mathsf{CSS}(X, \mathrm{H}_0; Z, \mathrm{G})$ *is a tri-orthogonal code. Then there exists a diagonal unitary operator* U *in the Clifford group* $\mathcal{K}^{(2)}$ *composed only of* cZ *and* S *such that*

$$\left|\overline{A^{\otimes k}}\right\rangle = UT_n\left|\overline{+^{\otimes k}}\right\rangle . \tag{12}$$

3 Distillation procedure

In this section, we describe the procedure we shall use to distill magic states. Let $\widetilde{\mathcal{C}} = [N, K]$ be a binary linear code defined by a parity check matrix $\widetilde{\mathrm{H}} \in \mathbb{F}_2^{(N-K)\times N}$. If the dual code $\widetilde{\mathcal{C}}^{\perp}$ is triply-even, we may construct a quantum code by puncturing the columns of $\widetilde{\mathrm{H}}$. For simplicity, suppose we puncture the first k bits of the space; if not, we can always permute the bits. We may then use Gaussian elimination to express the parity check matrix $\widetilde{\mathrm{H}}$ of the code $\widetilde{\mathcal{C}}$ in systematic form with respect to the puncture:

$$\widetilde{\mathrm{H}} = \begin{pmatrix} \mathbb{1}_k & \mathrm{H}_1 \\ 0 & \mathrm{H}_0 \end{pmatrix} , \tag{13}$$

where $\mathbb{1}_k$ represents the $k \times k$ identity matrix, and H_1, H_0 are matrices of dimension $k \times (N-k)$ and $(K-k) \times (N-k)$ respectively. The matrices H_1 and H_0 have rows whose weights are 1 and 0 mod 2 respectively. Since $\widetilde{\mathcal{C}}^{\perp}$ was a triply-even space by assumption, we obtain a tri-orthogonal matrix H as

$$\mathrm{H} = \begin{pmatrix} \mathrm{H}_1 \\ \mathrm{H}_0 \end{pmatrix} . \tag{14}$$

The resulting quantum code has block size $N-k$ and dimension k. Let $\mathcal{C}_P$ denote the punctured code obtained from $\mathcal{C}$ by removing the first k bits.

Let $\mathcal{T}$ represent the map $\mathcal{T}(\rho) = T\rho T^{\dagger}$, where the T gate is defined in eq. (1). For $p \in [0, 1]$, let $\mathcal{E}_p$ represent a single-qubit noise channel of strength p. In particular, we will be interested in channels whose action on a density operator ρ describing a single qubit is described by

1. the erasure channel $\mathcal{E}_p(\rho) = (1-p)\rho + p\left|\perp\right\rangle\left\langle\perp\right|$, where $\perp$ is a symbol denoting erasure, or

2. the dephasing channel $\mathcal{E}_p(\rho) = (1-p)\rho + pZ\rho Z$ [1].

We model the noisy T as the composition of ideal gate and the noise channel, i.e. $\mathcal{E}_p \circ \mathcal{T}$.

Distillation proceeds as follows:

1. We begin with the state $\eta_0 = \left|\overline{+^{\otimes k}}\right\rangle\left\langle\overline{+^{\otimes k}}\right|$, the encoded k-fold tensor product of the $|+\rangle$ state over $(N-k)$ physical qubits.

2. We apply an $(N-k)$-fold tensor product of (noisy) transversal T gates on η_0 to obtain η_1

$$\eta_1 = (\mathcal{E}_p \circ \mathcal{T})^{\otimes(N-k)}(\eta_0) .$$

[1] As noted by Bravyi and Haah, if the state is not already in this form, we can force it to be by applying the gates I and $e^{-i\pi/4}SX$ with probability 1/2 each.

3. We apply the Clifford gate guaranteed by lemma 10 to η_1 to obtain η_2

$$\begin{aligned}\eta_2 = U\eta_1 U^\dagger &= U\mathcal{E}_p\left(T_{N-k}\big|\overline{+^{\otimes k}}\big\rangle\big\langle\overline{+^{\otimes k}}\big|T_{N-k}^\dagger\right)U^\dagger \\ &= \mathcal{E}_p\left(UT_{N-k}\big|\overline{+^{\otimes k}}\big\rangle\big\langle\overline{+^{\otimes k}}\big|T_{N-k}^\dagger U^\dagger\right) \\ &= \mathcal{E}_p\left(\big|\overline{A^{\otimes k}}\big\rangle\big\langle\overline{A^{\otimes k}}\big|\right) \ .\end{aligned}$$

4. At this stage, we utilize Steane's error correction technique [40]. Upon measurement, we obtain a codeword of $\mathcal{C}_P$, and assume that the decoder has been provided the location of the punctures as side information. These locations together with the error arising from the noisy channel $\mathcal{E}_p$ can be regarded as a composite error channel on the code $\mathcal{C}$. The decoder can then proceed to run the decoder of $\mathcal{C}$ on the received codeword where it treats the punctured positions as having suffered erasure noise.

5. Having deduced the error E, we can now perform correction and run the encoder in reverse to obtain k copies of the magic state $|A\rangle$.

We note that our scheme uses Steane error correction at step 4 rather than measuring the X-stabilizers and post-selecting on the all $+1$-outcome. The polar code is designed to optimize its performance under SC decoding and therefore characterizing the code by its distance can be misleading. As mentioned in the introduction, the polar code outperforms the Reed-Muller code as measured by its error correction capacity even though both code families are constructed using the same encoding circuits. Therefore a better metric to study the performance of the polar code is to use its bit error probability under SC decoding. To compare the two schemes, note that the probability of successfully post-selecting a state scales as $O\left((1-p)^d\right)$, and therefore decreases as a function of the block-size (and the number of encoded qubits); the corresponding error rate falls as $O\left(p^d\right)$. On the other hand, decoding guarantees that every state obtained after error correction is part of the codespace whereas the error rate only falls as $O\left(\sqrt{p^d}\right)$. The probability of error in step 4 is thus upper-bounded by the probability of failure of the classical decoder associated with $\mathcal{C}$. Understanding the consequences for these trade-offs numerically are directions for future research.

4 Tri-orthogonal codes from polar codes

In this section, we demonstrate the existence of triply-even codes derived from polar codes. We supplement this with numerics which upper bound the size of the triply-even space and the probability of error associated with the decoding process for block sizes of interest.

Note that according to the prescription above, if we wish to construct a tri-orthogonal code from a code $\mathcal{C}$, we first require its dual $\mathcal{C}^\perp$ to be triply-even. The following series of simple lemmas establish constraints on a decreasing set I such that $\mathcal{C}(\mathrm{I})^\perp$ is triply-even.

Lemma 11. *If $\mathcal{V} \subseteq \mathbb{F}_2^N$ is triply-even is equivalent to stating that for any two vectors $u, v \in \mathcal{V}$, $u * v \in \mathcal{V}^\perp$.*

Proof. For $u, v, w \in \mathcal{V}$, $|u * v * w| = 0 \pmod 2$ and therefore, any product of the form $u * v$ is orthogonal to vectors in $\mathcal{V}$. The other direction also follows trivially. □

Lemma 12. *If $\mathrm{I}, \mathrm{J} \subseteq \mathcal{M}_n$ are two sets of monomials, then*

$$\mathcal{C}(\mathrm{I}) * \mathcal{C}(\mathrm{J}) = \mathcal{C}(\mathrm{I} \cdot \mathrm{J}) \ .$$

Proof. Let $\mathrm{I} = \{\alpha_1, ..., \alpha_{|\mathrm{I}|}\}$ and $\mathrm{J} = \{\beta_1, ..., \beta_{|\mathrm{J}|}\}$ denote the elements of these sets. Any vectors $v \in \mathcal{C}(\mathrm{I})$ and $w \in \mathcal{C}(\mathrm{J})$ can be expressed as $v = \sum_{i=1}^{|\mathrm{I}|} s_i \mathsf{ev}(\alpha_i)$ and $w = \sum_{j=1}^{|\mathrm{J}|} t_j \mathsf{ev}(\beta_j)$ for some constants $s_i, t_j \in \mathbb{F}_2$

for $i \in \mathrm{I}, j \in \mathrm{J}$. Since ev is a bijection and maps the star product of two vectors to the product of monomials, any element of the star product of the two is of the form

$$\sum_{i,j} r_{ij} \mathsf{ev}(\alpha_i) * \mathsf{ev}(\beta_j) = \sum_{i,j} r_{ij} \mathsf{ev}(\alpha_i \cdot \beta_j) \ ,$$

for some constants $r_{ij} \in \mathbb{F}_2$ for $i \in \mathrm{I}, j \in \mathrm{J}$. The desired result follows. □

Lemma 13. *Let $\mathrm{I} \subset \mathcal{M}_n$ be decreasing and $\mathcal{C}(\mathrm{I})$ be the corresponding decreasing monomial code of dimension K. The space $\mathcal{C}(\mathrm{I})^\perp = \mathcal{C}(\mathcal{M}_n \setminus \check{\mathrm{I}})$ is tri-orthogonal if and only if*

$$(\mathcal{M}_n \setminus \check{\mathrm{I}}) \cdot (\mathcal{M}_n \setminus \check{\mathrm{I}}) \subseteq \mathrm{I} \ .$$

Proof. From lemma 11, the requirement that the code $\mathcal{C}^\perp = \mathcal{C}(\mathcal{M}_n \setminus \check{\mathrm{I}})$ be triply-even is equivalent to

$$\mathcal{C}(\mathcal{M}_n \setminus \check{\mathrm{I}}) * \mathcal{C}(\mathcal{M}_n \setminus \check{\mathrm{I}}) \subseteq \mathcal{C}(\mathrm{I}) \ . \tag{15}$$

From lemma 12, it follows that this is equivalent to

$$(\mathcal{M}_n \setminus \check{\mathrm{I}}) \cdot (\mathcal{M}_n \setminus \check{\mathrm{I}}) \subseteq \mathrm{I} \ . \tag{16}$$

□

These results imply the following corollary which is straightforward to check numerically and doing so gives us the size of the dual code $\mathcal{C}^\perp$.

Corollary 14. *Let $\mathrm{I} \in \mathcal{M}_n$ be a decreasing set of monomials. Finding the smallest code $\mathcal{C}(\mathrm{I})$ such that $\mathcal{C}^\perp$ is triply-even corresponds to finding the smallest set I for which the condition*

$$(\mathcal{M}_n \setminus \check{\mathrm{I}}) \cdot (\mathcal{M}_n \setminus \check{\mathrm{I}}) \cap (\mathcal{M}_n \setminus \mathrm{I}) = \varnothing \tag{17}$$

is still true.

Proof. Assume that $f \in \mathcal{M}_n \setminus \check{\mathrm{I}}$ but $f \cdot f \notin \mathrm{I}$. This implies that $f \cdot f \in \mathcal{M}_n \setminus \mathrm{I}$, resulting in the condition 17 above. Therefore the smallest code $\mathcal{C}(\mathrm{I})$ such that $\mathcal{C}^\perp$ is triply-even is the smallest code for which the condition

$$(\mathcal{M}_n \setminus \check{\mathrm{I}}) \cdot (\mathcal{M}_n \setminus \check{\mathrm{I}}) \cap (\mathcal{M}_n \setminus \mathrm{I}) = \varnothing$$

is still true. □

We wish to design a polar code $\mathcal{C}$ that will be subject to N copies of a noise channel $\mathsf{W} : \mathbb{F}_2 \to \mathcal{Y}$ such that its dual $\mathcal{C}^\perp$ is triply-even. To do so, we first compute the Bhattacharyya parameters corresponding to this channel as explained in subsection 2.1. We then order the indices $a \in \{0,1\}^N$ according to this parameter, or equivalently, we order the monomials that corresponds to each index a. Given a rate $R \in [0,1]$, we construct the polar code $\mathcal{C}$ using the RN best channels chosen according to their performance. The threshold $\epsilon \in [0,1]$ which is an upper bound on the bit error rate is the maximum Bhattacharyya parameter for the synthetic channels in $\mathcal{C}$.

These synthetic channels are in one-to-one correspondence with monomials as described in subsection 2.2 and so equivalently, this procedure yields a set of monomials I of size RN. We then use corollary 14 to verify whether the set $\mathcal{M}_n \setminus \check{\mathrm{I}}$ satisfies the desired product relation. Our goal is to construct the smallest set I satisfying this corollary as this will result in the largest code $\mathcal{C}^\perp$, which in turn will maximize the rate of the resulting quantum code.

We remark that this method may not be optimal. It may be entirely possible that the onerous synthetic channels that violate the triply-even condition are not the poorly performing channels. We leave the task of optimizing this algorithm as an objective for future research.

4.1 Dimension of triply-even polar codes

Recall that polar codes are channel specific – they have to be designed for a particular noise channel. We first discuss the erasure channel as it is the simplest case to study the polar code. As explained above, we can estimate the dimension of the channel numerically. Figure 1 shows how the rate of the dual code varies as a function of the log of the block size N for erasure channels whose probability of erasure are 0.01, 0.02 and 0.05 respectively. For an erasure rate of $p = 0.01$, the log-log plot and the corresponding line of best fit

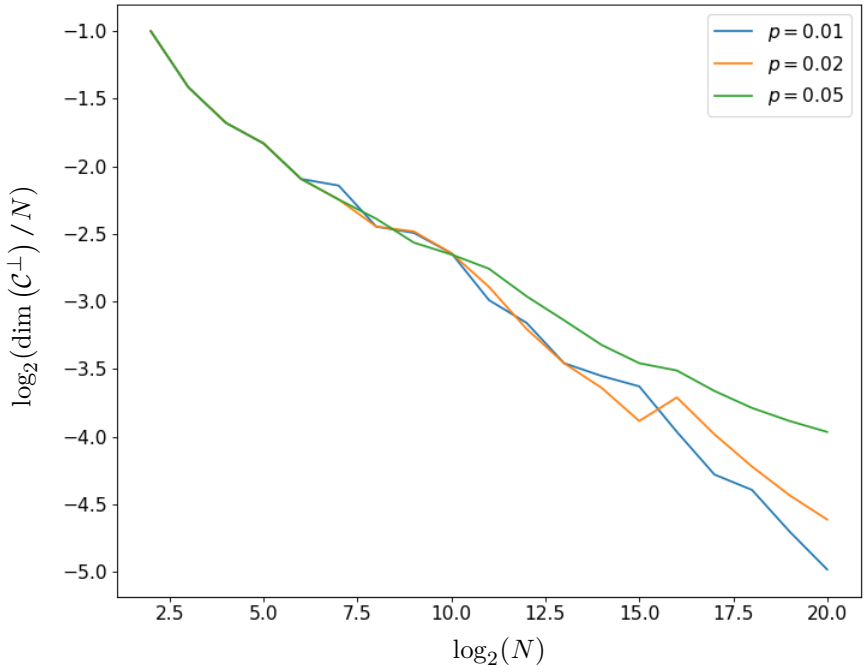

Figure 1: Log-log plot of rate $\dim\left(\mathcal{C}^{\perp}\right)/N$ of the dual code vs the block-size N for erasure channels with erasure probabilities 0.01, 0.02 and 0.05.

indicate that the rate of the code $\mathcal{C}^{\perp}$ scales roughly as

$$\dim\left(\mathcal{C}^{\perp}\right)/N = O(N^{-0.2}) \ ,$$

or equivalently, that the dimension of the code scales as

$$\dim\left(\mathcal{C}^{\perp}\right) = O(N^{0.8}) \ .$$

We also add that the rates of the codes for larger values of p are better but as we shall soon see, this comes at the cost of error tolerance.

We now proceed to study the binary symmetric channel. The complexity of designing the polar code as per Arikan's formulation using the Bhattacharyya parameters grows exponentially in the block size. We instead use Monte Carlo sampling techniques to estimate the Bhattacharyya parameters, the drawback of which is that we require a very large number of samples before we see convergence. For this reason, the range of block sizes we have explored is not as extensive as that of the erasure channel, and stops at $2^{16} = 65536$ qubits. Fig. 2 is a log-log plot of the rate $\dim\left(\mathcal{C}^{\perp}\right)/N$ of the dual code versus the block-size N. Using a

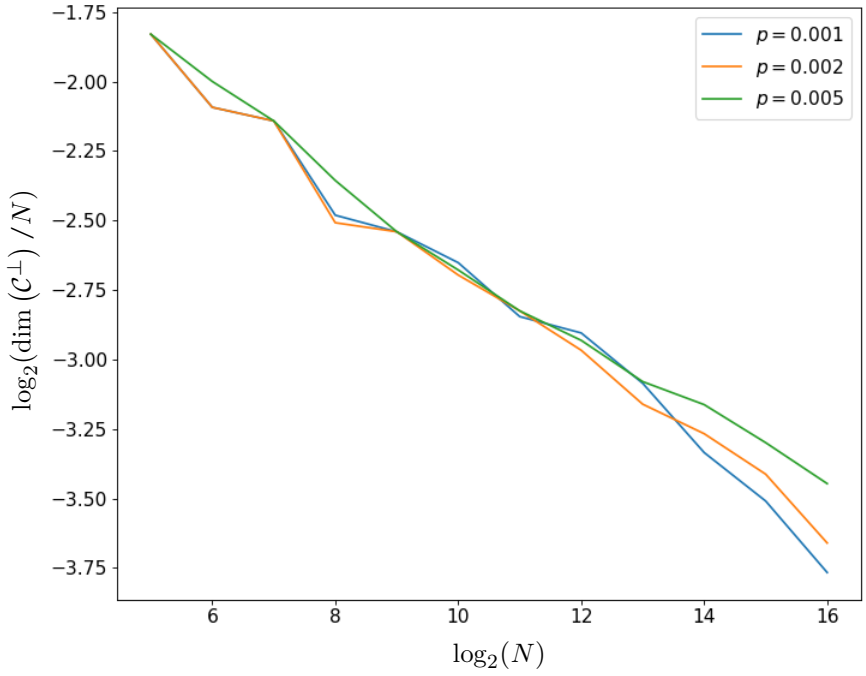

Figure 2: Log-log plot of rate $\dim\left(\mathcal{C}^{\perp}\right)/N$ of the dual code vs the block-size N for binary symmetric channels at transition probabilities 10^{-3}, 2×10^{-3} and 5×10^{-3}.

line of best fit, we find that for noise rate $p = 0.001$, the dimension of $\mathcal{C}^{\perp}$ versus the block size N scales roughly as $O(N^{0.84})$.

This can be contrasted to the scheme by Hastings and Haah who used punctured Reed-Muller codes to achieve $\mathcal{C}^{\perp} = O(N^{0.91})$. Note however that this dimension does not tell the whole story as this is only an upper bound on the dimension of the tri-orthogonal quantum code. More importantly, there is a tradeoff in decoding complexity and performance when using Reed-Muller codes and successive decoders. The better rate and lower encoding complexity make the polar code more favorable.

The polar code construction is only guaranteed to perform well if the rate of the code $\mathcal{C}$ is below the capacity of the channel it is designed for. For the erasure channel W_p with erasure probability $p \in [0,1]$, the capacity is $(1-p) \in [0,1]$ and for the binary symmetric channel V_q with transition probability $q \in [0,1/2]$, the capacity is $1 - h_2(q) \in [0,1]$, where $h_2(\cdot)$ is the binary entropy [41]. However, the triply-even constraint implies that the code $\mathcal{C}(\mathrm{I})$ will eventually encompass the entire space and exceed the maximum size of the polar code for a given noise rate that is guaranteed to perform well. At that point, this construction is no longer valid as this will mean that the SC decoding can no longer be guaranteed to work.

4.1.1 Error rates for the binary erasure channel

In this section, we shall prove that there exist punctures that do not compromise the performance of the code. The polar code is not designed to maximize the distance, but rather to minimize the probability of decoding error under SC decoding. A punctured polar code can be seen as a polar code that has suffered erasure errors – although the punctured bits are not transmitted, the decoder can replace the punctured positions with erasure symbols given the locations of the puncture as side information. If the noise channel that the code encounters is also an erasure channel, then these two processes – puncturing and noise – can

be seen as the composition of two erasure channels. This can be modeled easily thanks to the following observation whose proof is simple and therefore omitted.

Lemma 15. *The composition of two erasure channels W_p and W_q with erasure probabilities p and q respectively is an erasure channel W_r with erasure probability $r = p + q - pq$.*

Let $\mathcal{C}$ be a polar code whose dual is triply-even and is designed for the erasure channel W_r. We denote the threshold of $\mathcal{C}$ by $\epsilon \in [0, 1]$. For $p, q \in [0, 1]$ such that $p + q - pq = r$, we puncture our polar code $\mathcal{C}$ randomly using an erasure channel W_q and it is then subject to erasure noise W_p.

To show that there exists a good punctured code, we employ a derandomization argument.

Lemma 16. *For $\epsilon_0 > \epsilon$, there exists a $\delta \in [0, 1]$ such that we can choose with probability δ a punctured polar code $\widetilde{\mathcal{C}}_{\epsilon_0}$ whose bit error rate is upper bounded by ϵ_0 against erasure noise W_p for $p < r$.*

Proof. For $q \in [0, 1]$, let W_q denote the erasure channel that creates the puncture. We first apply the channel W_q^N which serves to create the punctured code. Then we apply W_p^N which is the noise that the punctured code is subject to. According to lemma 15, the effective channel that $\mathcal{C}$ is subject to is W_r^N, where $r = p + q - pq$. Upon transmitting the message $u \in \mathbb{F}_2^N$ across this channel, this could result in erasure patterns that we denote $P, Q \in \mathbb{F}_2^N$ where location a has suffered an erasure error if and only if $(P \vee Q)_a = 1$. Let $\Pr\{u_a \neq \hat{u}_a | P \vee Q\}$ be the probability that SC decoding results in a decoding error for location a.

Since the threshold of $\mathcal{C}$ is ϵ, we can upper bound the bit error rate as

$$\mathbb{E}_Q \mathbb{E}_P \Pr\{u_a \neq \hat{u}_a | P \vee Q\} \leq \epsilon \ , \tag{18}$$

where $\mathbb{E}_Q$ and $\mathbb{E}_P$ denote the expectation value over random erasure patterns $Q, P \in \mathbb{F}_2^N$.

For $\delta \in [0, 1]$, we may now apply a Markov inequality to upperbound the probability of picking a 'bad' erasure pattern Q

$$\Pr_Q \left\{ \mathbb{E}_P \Pr\{u_a \neq \hat{u}_a | P \vee Q\} \geq \epsilon_0 \right\} \leq \delta \ , \tag{19}$$

where we define the threshold $\delta = \epsilon / \epsilon_0$. Hence for $\epsilon_0 > \epsilon$, there exist punctures P such that the probability of failure for any bit is upper bounded by ϵ_0. □

Let $\mathcal{C}$ be the smallest polar code such that $\mathcal{C}^\perp$ is triply-even, and let ϵ be its threshold. In fig. 3 below, we plot the corresponding log-likelihood ratios (LLRs) vs the log of the block size. In this case, the LLR corresponding to $\mathcal{C}$ is $\log_2((1 - \epsilon)/\epsilon)$.

We find that the probability of decoding error drops significantly as the block-size increases. This suggests that for block sizes of interest, it might suffice to increase the block size of the polar code rather than perform concatenation as in the Bravyi-Haah distillation scheme. We can see that the code designed for $p = 0.01$ has an error rate several orders of magnitude better for block sizes of interest. Thus for a minor tradeoff in the dimension with respect to codes designed for $p = 0.02$ and $p = 0.05$, the code designed for $p = 0.01$ sees a tremendous benefit with respect to the error rate. We can see that for a block size of $2^{18} = 262,144$ at erasure error rate 0.01, we can achieve a bit error rate of roughly $2^{-90} \approx 8 \times 10^{-28}$.

4.1.2 Error rates for the binary symmetric channel

In the event that the noise channel that the quantum code is subject to is a dephasing channel, then we must deal with a composition of a binary erasure channel (which creates the puncture) and a binary symmetric channel (which is the noise). Before doing so, it will be convenient to introduce the notion of channel

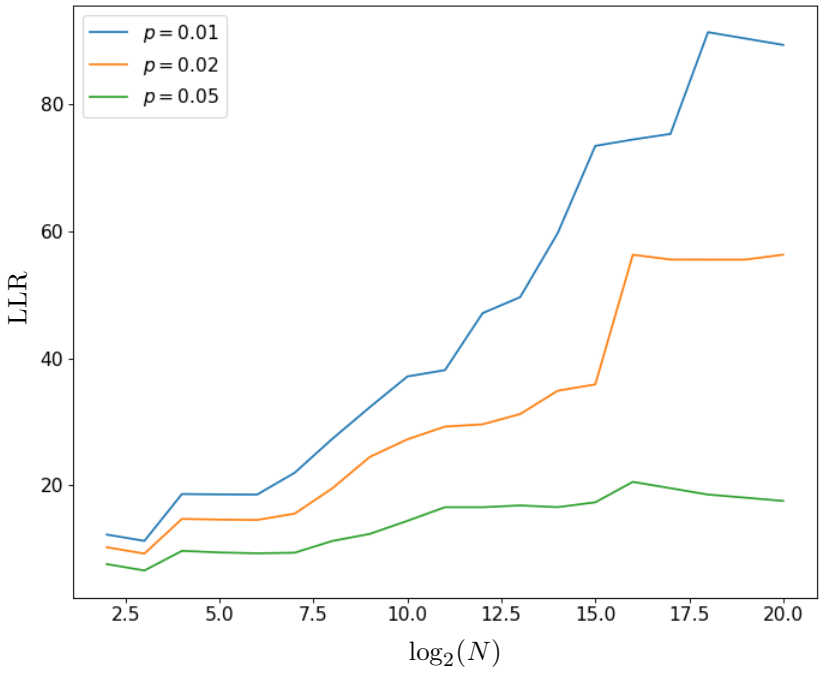

Figure 3: Variation of the log-likelihood ratios (LLRs) vs the log of the block-size for erasure probabilities $p = 0.01$, $p = 0.02$ and $p = 0.05$.

degradability as in [1]. For some sets of alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, we say that a channel $\mathsf{U} : \mathcal{X} \to \mathcal{Y}$ can be *degraded* to a channel $\widetilde{\mathsf{U}} : \mathcal{X} \to \mathcal{Z}$ if there exists a channel $\mathsf{Q} : \mathcal{Y} \to \mathcal{Z}$ such that

$$\widetilde{\mathsf{U}}(z|x) = \sum_{y \in \mathcal{Y}} \mathsf{Q}(z|y)\mathsf{U}(y|x) \ .$$

If such a channel Q exists, this is denoted $\widetilde{\mathsf{U}} \preccurlyeq \mathsf{U}$. Equivalently, we can say that U can be degraded to U' if there exists a channel Q that uses U as a sub-routine to simulate $\widetilde{\mathsf{U}}$.

We use the following lemma, each of whose claims has been proved earlier.

Lemma 17. *For channels $\mathsf{U} : \mathcal{X} \to \mathcal{Y}$ and $\widetilde{\mathsf{U}} : \mathcal{X} \to \mathcal{Y}$,*

1. *([42] (p207)) the reliability of U is at least as good as that of $\widetilde{\mathsf{U}}$, i.e.*

$$\widetilde{\mathsf{U}} \preccurlyeq \mathsf{U} \implies \mathcal{B}(\mathsf{U}) \leq \mathcal{B}(\widetilde{\mathsf{U}}) \ ;$$

2. *([43] (lemma 4.7)) for all block sizes N, the synthetic channels inherit this relation, i.e. for $a \in [N]$,*

$$\widetilde{\mathsf{U}} \preccurlyeq \mathsf{U} \implies \widetilde{\mathsf{U}}^{(a)} \preccurlyeq \mathsf{U}^{(a)} \ .$$

In other words, if U can be degraded to U', then U is at least as reliable as U'. Using the notion of degradability, we can simplify the analysis for the binary symmetric channel by degrading the erasure channel using the following lemmas.

Lemma 18. *Let $p \in [0, 1]$ be some erasure probability.*

1. *The erasure channel W_p of erasure probability p is degradable to a binary symmetric channel of transition probability $p/2$, i.e.*

$$\mathsf{V}_{p/2} \preccurlyeq \mathsf{W}_p \ .$$

2. *The composition of binary symmetric channels V_a and V_b is a binary symmetric channel V_r where $r = p + q - 2pq$.*

Proof. We shall prove each of these statements in turn.

1. Consider transmitting a bit across the binary erasure channel W_p of erasure probability p. With probability p, the bit is erased and we may replace it by a random symbol 0 or 1. Hence with probability p, the effective channel $\mathcal{Z}$ is a binary symmetric channel with probability $1/2$ and is the ideal channel otherwise. For $x \in \mathbb{F}_2$, we have

$$\begin{aligned}\mathcal{Z}(x) &= (1-p)\mathsf{V}_0(x) + p\mathsf{V}_{1/2}(x) \\ &= (1-p)[x] + \frac{p}{2}[x] + \frac{p}{2}[\bar{x}] \\ &= \left(1 - \frac{p}{2}\right)[x] + \left(\frac{p}{2}\right)[\bar{x}] \ .\end{aligned}$$

   Thus $\mathcal{Z}$ is equivalent to the binary symmetric channel V_r, where $r = p/2$.

2. For $x \in \mathbb{F}_2$, the action of the composition $\mathsf{V}_a \circ \mathsf{V}_b$ is

$$\begin{aligned}\left(\mathsf{V}_a \circ \mathsf{V}_b\right)(x) &= \mathsf{V}_a\left((1-b)[x] + b[\bar{x}]\right) \\ &= \left((1-a)(1-b) + ab\right)[x] + \left(a(1-b) + (1-a)b\right)[\bar{x}] \\ &= \left(1 - a - b + 2ab\right)[x] + \left(a + b - 2ab\right)[\bar{x}] \ .\end{aligned}$$

□

Of course, estimating the probability of failure by degrading the erasure channel to a binary symmetric channel only yields a lower bound. This is because we know the erasure locations when we transmit across the BEC but this is no longer true when we code for the binary symmetric channel.

Let $\mathcal{C}$ be a polar code whose dual is triply-even and is designed for the composition of the erasure channel W_q and binary symmetric channel V_p. We denote the threshold of $\mathcal{C}$ by $\epsilon \in [0,1]$. For $p, q \in [0,1]$ such that $p + q/2 - pq = r$, we puncture our polar code $\mathcal{C}$ randomly using an erasure channel W_q and it is then subject to the binary symmetric channel V_p.

As was done for the erasure channel, to show that there exists a good punctured code, we employ a derandomization argument.

Lemma 19. *For $\epsilon_0 > \epsilon$, there exists a $\delta \in [0,1]$ such that we can choose with probability δ a punctured polar code $\widetilde{\mathcal{C}}_{\epsilon_0}$ whose bit error rate is upper bounded by ϵ_0 against erasure noise W_p for $p < r$.*

Proof. The code $\mathcal{C}$ is designed for the composition $\mathcal{Z} := \mathsf{W}_q \circ \mathsf{V}_p$ of the erasure channel W_q and the binary symmetric channel V_p. The binary erasure channel W_q can be degraded to a binary symmetric channel $\mathsf{V}_{q/2}$ according to lemma 18. Noting that the composition of the channels V_p and $\mathsf{V}_{q/2}$ is yet again a binary symmetric channel V_r, where $r = p + q/2 - pq$, we have $\mathsf{V}_r \preccurlyeq \mathcal{Z}$. It follows then from lemma 17 that the synthetic channels obtained by designing polar codes for $\mathcal{Z}$ can be degraded to those synthetic channels obtained by designing polar codes for V_r.

The proof now follows similarly to the case of the erasure channel we have dealt with.

Transmitting the message $u \in \mathbb{F}_2^N$ across V_r could result in error patterns that we denote $P, Q \in \mathbb{F}_2^N$ where location a has suffered an erasure error if and only if $(P+Q)_a = 1$. Let $\Pr\{u_a \neq \hat{u}_a | P + Q\}$ be the probability that SC decoding results in a decoding error for location a.

Since the threshold of $\mathcal{C}$ is ϵ, we can upper bound the bit error rate as

$$\mathbb{E}_Q \mathbb{E}_P \Pr\{u_a \neq \hat{u}_a | P + Q\} \leq \epsilon \ , \tag{20}$$

where $\mathbb{E}_Q$ and $\mathbb{E}_P$ denote the expectation value over random error patterns Q, P.

For $\delta \in [0, 1]$, we may now apply a Markov inequality to upperbound the probability of picking a 'bad' erasure pattern Q

$$\Pr_Q \{\mathbb{E}_P \Pr\{u_a \neq \hat{u}_a | P + Q\} \geq \epsilon_0\} \leq \delta \ , \tag{21}$$

where we define the threshold $\delta = \epsilon/\epsilon_0$. Hence for $\epsilon_0 > \epsilon$, there exist punctures P such that the bit error rate for any bit is upper bounded by ϵ_0. □

Let $\mathcal{C}$ be the smallest polar code such that $\mathcal{C}^\perp$ is triply-even, and let ϵ be its threshold. In figure 4 below, we plot $-\log_2(\epsilon)$ versus the log of the block size of the code $\mathcal{C}$ for noise rates 0.001, 0.002 and 0.005. For the noise rates we are studying, we expect the performance of the polar codes designed for the binary symmetric channel to be worse than those designed for the binary erasure channel. We can see that for block sizes of $2^{16} = 65,536$ for a noise rate of 0.001, we can achieve a bit error rate of roughly $2^{-46} \approx 7 \times 10^{-15}$.

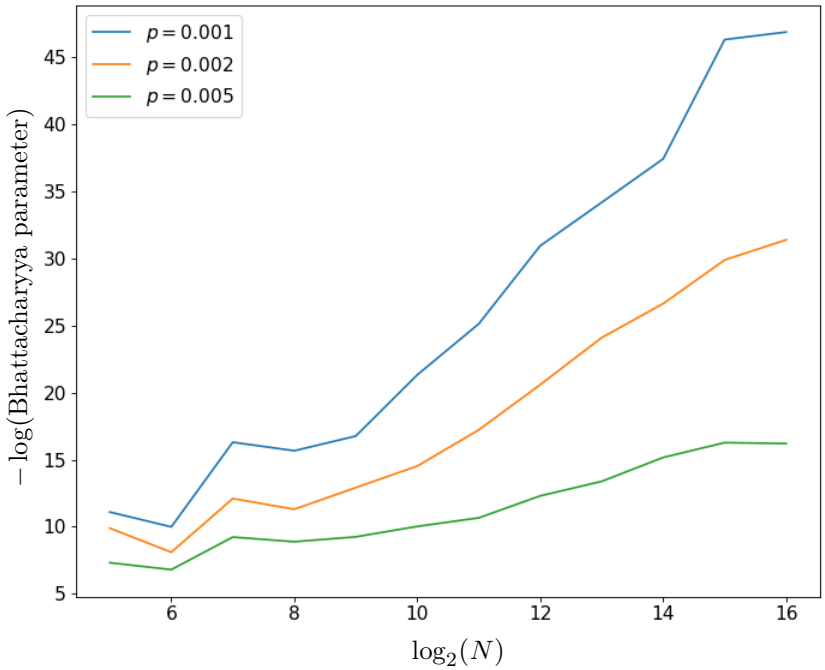

Figure 4: Log-log plot of the Bhattacharyya parameters vs the block-size for transition probabilities $p = 0.001$, $p = 0.002$ and $p = 0.005$.

5 Conclusions

We have demonstrated how to use polar codes to construct tri-orthogonal codes, which in turn can be used for magic state distillation using the framework of Bravyi and Haah. We introduced new theoretical tools

from recent advances in classical coding theory which we hope will help in the broader study of quantum error correcting codes. In addition, using polar codes allows us to use the low complexity successive cancellation decoder which functions in $O(N \log N)$ time for a polar code of block size N. By puncturing these codes, we can achieve a growing dimension for the number of encoded logical qubits. This decoder is known to have a better error correction capacity than decoders for Reed-Muller codes. The dimension of these triply-even codes is shown to scale like $O(N^{0.8})$ for the binary erasure channel at noise rate 0.01 and $O(N^{0.84})$ for the binary symmetric channel at noise rate 0.001. Finally it was shown that we can upper bound the probability of failure of decoding a bit using the Bhattacharyya parameters for the corresponding classical codes. The error probability for the triply-even codes can drop to roughly 8×10^{-28} for the erasure channel at a block size of $2^{20} = 262,144$ and 7×10^{-15} for the dephasing channel at a block size of $2^{16} = 65,536$.

Future directions: It is well known that polar codes equipped with the successive cancellation decoder require large block-lengths to be effective. In order to address this problem, Tal and Vardy have devised a successive cancellation list decoder [44]. Together with some pre-encoding with cyclic repetition codes, the performance of these codes is significantly improved even for smaller block sizes. For practical implementations, it would be interesting to see how these ideas can be used to improve the performance of polar codes for magic state distillation. It would also be interesting to perform numerical simulations to see how the performance of punctured polar codes directly compares to punctured Reed-Muller codes.

6 Acknowledgements

We would like to thank David Poulin for many helpful discussions. In particular, we would like to thank him for drawing our attention to the argument in section 3 contrasting error correction and post-selection. We would also like to thank Colin Trout for detailed feedback on an earlier draft of this work, and Pavithran Iyer and Maxime Tremblay for many helpful comments. A.K. would like to thank the MITACS organization for the Globalink award which facilitated his visit to Inria, Paris and Inria, Paris for their hospitality during his visit. A.K. also acknowledges support from the Fonds de Recherche du Québec - Nature et Technologies (FRQNT) via the B2X scholarship for doctoral candidates. Computations were made on the supercomputers managed by Calcul Québec and Compute Canada. The operation of these supercomputers is funded by the Canada Foundation for Innovation (CFI), the Ministère de l'Economie, de la Science et de l'Innovation du Québec (MESI) and the FRQNT. JPT acknowledges the support of the European Union and the French Agence Nationale de la Recherche through the QCDA project.

References

[1] M. Bardet, V. Dragoi, A. Otmani, and J.-P. Tillich. Algebraic properties of polar codes from a new polynomial formalism. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 230–234. IEEE, 2016.

[2] S. Bravyi and J. Haah. Magic-state distillation with low overhead. *Physical Review A*, 86(5):052329, 2012.

[3] P. W. Shor. Fault-tolerant quantum computation. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 56–65. IEEE, 1996.

[4] D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 176–188. ACM, 1997.

[5] A. Y. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[6] E. Knill, R. Laflamme, and W. H. Zurek. Resilient quantum computation: error models and thresholds. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 365–384. The Royal Society, 1998.

[7] S. Bravyi and A. Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316, 2005.

[8] C. Jones. Multilevel distillation of magic states for quantum computing. *Physical Review A*, 87(4):042305, 2013.

[9] J. O'Gorman and E. T. Campbell. Quantum computation with realistic magic-state factories. *Physical Review A*, 95(3):032338, 2017.

[10] M. B. Hastings and J. Haah. Distillation with sublogarithmic overhead. *Physical Review Letters*, 120(5):050504, 2018.

[11] J. Haah and M. B. Hastings. Codes and protocols for distilling t, controlled-s, and toffoli gates. *arXiv preprint arXiv:1709.02832*, 2017.

[12] E. Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, 2009.

[13] I. Dumer. Recursive decoding of reed-muller codes. *arXiv preprint arXiv:1703.05303*, 2017.

[14] E. Arikan. A survey of reed-muller codes from polar coding perspective. In *Information Theory (ITW 2010, Cairo), 2010 IEEE Information Theory Workshop on*, pages 1–5. IEEE, 2010.

[15] E. Arikan. A performance comparison of polar codes and reed-muller codes. *IEEE Communications Letters*, 12(6), 2008.

[16] J. M. Renes, F. Dupuis, and R. Renner. Efficient polar coding of quantum information. *Physical Review Letters*, 109(5):050504, 2012.

[17] J. M. Renes and M. M. Wilde. Polar codes for private and quantum communication over arbitrary channels. *IEEE Transactions on Information Theory*, 60(6):3090–3103, 2014.

[18] J. M. Renes, D. Sutter, F. Dupuis, and R. Renner. Efficient quantum polar codes requiring no preshared entanglement. *IEEE Transactions on Information Theory*, 61(11):6395–6414, 2015.

[19] M. M. Wilde and J. M. Renes. Quantum polar codes for arbitrary channels. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 334–338. IEEE, 2012.

[20] M. M. Wilde and S. Guha. Polar codes for classical-quantum channels. *IEEE Transactions on Information Theory*, 59(2):1175–1187, 2013.

[21] M. M. Wilde and S. Guha. Polar codes for degradable quantum channels. *IEEE Transactions on Information Theory*, 59(7):4718–4729, 2013.

[22] M. M. Wilde, O. Landon-Cardinal, and P. Hayden. Towards efficient decoding of classical-quantum polar codes. *arXiv preprint arXiv:1302.0398*, 2013.

[23] B. W. Reichardt. Quantum universality from magic states distillation applied to css codes. *Quantum Information Processing*, 4(3):251–264, 2005.

[24] A. M. Meier, B. Eastin, and E. Knill. Magic-state distillation with the four-qubit code. *arXiv preprint arXiv:1204.4221*, 2012.

[25] J. Haah, M. B. Hastings, D. Poulin, and D. Wecker. Magic state distillation at intermediate size. *arXiv preprint arXiv:1709.02789*, 2017.

[26] J. Haah, M. B. Hastings, D. Poulin, and D. Wecker. Magic state distillation with low space overhead and optimal asymptotic input count. *Quantum*, 1:31, 2017.

[27] A. J. Landahl and C. Cesare. Complex instruction set computing architecture for performing accurate quantum z rotations with less magic. *arXiv preprint arXiv:1302.3240*, 2013.

[28] G. Duclos-Cianci and D. Poulin. Reducing the quantum-computing overhead with complex gate distillation. *Physical Review A*, 91(4):042315, 2015.

[29] E. T. Campbell and M. Howard. Unifying gate synthesis and magic state distillation. *Physical Review Letters*, 118(6):060501, 2017.

[30] E. T. Campbell and M. Howard. Unified framework for magic state distillation and multiqubit gate synthesis with reduced resource cost. *Physical Review A*, 95(2):022316, 2017.

[31] H. Anwar, E. T. Campbell, and D. E. Browne. Qutrit magic state distillation. *New Journal of Physics*, 14(6):063006, 2012.

[32] E. T. Campbell, H. Anwar, and D. E. Browne. Magic-state distillation in all prime dimensions using quantum reed-muller codes. *Physical Review X*, 2(4):041021, 2012.

[33] E. T. Campbell. Enhanced fault-tolerant quantum computing in d-level systems. *Physical Review Letters*, 113(23):230501, 2014.

[34] A. Krishna and J.-P. Tillich. Towards low overhead magic state distillation. *arXiv preprint arXiv:1811.08461*, 2018.

[35] R. Mori and T. Tanaka. Performance of polar codes with the construction using density evolution. *IEEE Communications Letters*, 13(7), 2009.

[36] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar. On the construction of polar codes. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 11–15. IEEE, 2011.

[37] I. Tal and A. Vardy. How to construct polar codes. *IEEE Transactions on Information Theory*, 59(10):6562–6582, 2013.

[38] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Physical Review A*, 54(2):1098, 1996.

[39] A. Steane. Multiple-particle interference and quantum error correction. *Proc. R. Soc. Lond. A*, 452(1954):2551–2577, 1996.

[40] A. M. Steane. Active stabilization, quantum computation, and quantum state synthesis. *Physical Review Letters*, 78(11):2252, 1997.

[41] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[42] T. Richardson and R. Urbanke. *Modern coding theory*. Cambridge university press, 2008.

[43] K. Kobara. Code-based public-key cryptosystems and their applications. In *International Conference on Information Theoretic Security*, pages 45–55. Springer, 2009.

[44] I. Tal and A. Vardy. List decoding of polar codes. *IEEE Transactions on Information Theory*, 61(5):2213–2226, 2015.